

# Mixed and componentwise condition numbers for a linear function of the solution of the total least squares problem

Huai-An Diao<sup>a,\*</sup>, Yang Sun<sup>a</sup>

<sup>a</sup>*School of Mathematics and Statistics, Northeast Normal University, No. 5268 Renmin Street, Chang Chun 130024, P.R. of China.*

---

## Abstract

In this paper, we consider the mixed and componentwise condition numbers for a linear function of the solution to the total least squares (TLS) problem. We derive the explicit expressions of the mixed and componentwise condition numbers through the dual techniques. The sharp upper bounds for the derived mixed and componentwise condition numbers are obtained. For the structured TLS problem, we consider the structured perturbation analysis and obtain the corresponding expressions of the mixed and componentwise condition numbers. We prove that the structured ones are smaller than their corresponding unstructured ones based on the derived expressions. Moreover, we point out that the new derived expressions can recover the previous results on the condition analysis for the TLS problem. The numerical examples show that the derived condition numbers can give sharp perturbation bounds, on the other hand normwise condition numbers can severely overestimate the relative errors because normwise condition numbers ignore the data sparsity and scaling. Meanwhile, from the observations of numerical examples, it is more suitable to adopt structured condition numbers to measure the conditioning for the structured TLS problem.

*Keywords:* Total least squares problem, componentwise perturbation, condition number, adjoint operator, structured perturbation.

*2010 MSC:* 15A09, 15A12, 65F35

---

## 1. Introduction

For a given over-determined set of  $m$  linear equations  $Ax \approx b$  in  $x \in \mathbb{R}^n$ , the total least squares (TLS) problem [1, 2, 3] is defined by

$$\begin{aligned} & \text{minimize} && \|[A, b] - [\hat{A}, \hat{b}]\|_F \\ & \text{subject to} && \hat{b} \in \mathcal{R}(\hat{A}), [\hat{A}, \hat{b}] \in \mathbb{R}^{m \times (n+1)}, \end{aligned} \tag{1.1}$$

---

\*Corresponding author

<sup>1</sup>Email address: hadiao@nenu.edu.cn, hadiao78@yahoo.com (H.A. Diao), suny235@nenu.edu.cn (Y. Sun).

where  $A \in \mathbb{R}^{m \times n}$ ,  $b \in \mathbb{R}^m$ ,  $R(A)$  denotes the range space of the matrix  $A$  and  $\|\cdot\|_F$  is the Frobenius norm. Let  $[\hat{A}, \hat{b}]$  is a minimizer of (1.1), then any  $x$  satisfying  $\hat{A}x = \hat{b}$  is called a TLS solution and  $[\widehat{\Delta A}, \widehat{\Delta b}] = [A, b] - [\hat{A}, \hat{b}]$  is the corresponding TLS correction. In order to guarantee the existence and uniqueness of the TLS solution to (1.1), the *genericity* condition (2.2) of the TLS problem was first introduced in [1]. In this paper, we always assume that the genericity condition holds (for more information about the *nongeneric* problem, see, e.g.[3]). The TLS problem has many applications in computer vision, image reconstruction, speech and audio processing, modal and spectral analysis, linear system theory, and system identification, etc; see the review paper [4] of TLS for more details.

Now we first review numerical methods for TLS. When the size of TLS problem is small or medium, a classical direct solver based on the full singular value decomposition (SVD) of the augmented matrix  $[A, b]$  can be adopted; see [3, Algorithm 3.1]. The solution in the generic case can be obtained from the right singular vector corresponding to the smallest singular value of  $[A, b]$ . The improved version of the SVD-based direct method can be implemented by using the partial SVD of  $[A, b]$  to compute in an efficient and reliable way a basis of the left and/or right singular subspace of a matrix associated with its smallest singular values; see [3, Chapter 4] for more details. The iterative method combining the Rayleigh quotient iteration and preconditioned conjugate gradient method was proposed in [5] for the large-scale and sparse TLS problem. A lot of researchers had paid attentions to the numerical solver for the large-scale structured TLS problem; see the papers [6, 7, 8, 9, 10, 11].

In sensitivity analysis, the condition number is considered as a fundamental tool since they describes the worst-case sensitivity of the solution to a problem with respect to small perturbations in the input data. The problem with a large condition number is called an *ill-posed* problem (cf. [12]). Since the 1980's, there had been some papers related to the perturbation analysis for the TLS problem; see [13, 1, 14] and the references therein. As far as we know, the general normwise condition numbers were studied by Rice [15], which measure errors for both input and output data by means of norm. However, when the data is badly-scaled or sparse, normwise condition numbers may allow large relative perturbations on small entries and give over-estimated perturbation bounds. To overcome the shortcoming of normwise condition numbers, componentwise perturbation analysis has been extensively studied for many classical problems in matrix computation; see the comprehensive survey [16] and the references therein. Because rounding errors for the data in the floating point system are measured componentwisely, it is more reasonable to adopt componentwise perturbation analysis and more sharper bounds can be obtained through componentwise perturbation analysis. In fact, most error bounds in LAPACK [17] are based on componentwise perturbation analysis. In componentwise perturbation analysis, two types of condition numbers, described as mixed and componentwise, were proposed; see [18, 19, 20, 21] for details.

Under the genericity condition, normwise condition numbers for the TLS problem had been studied in [22, 23]. Specifically, the explicit expressions, their lower and upper bounds, replying on the SVDs of the matrix  $A$  and/or the augmented matrix  $[A, b]$ , were derived. However, as stated in the previous paragraph, when the data is sparse or badly scaled, normwise condition numbers may heavily over-estimate the conditioning of the TLS problem.

Thus it is necessary to consider the conditioning of the TLS problem using componentwise perturbation analysis, which had been done in [24]. As shown [25, Example 1], there are big differences between normwise condition numbers and mixed/componentwise condition numbers, which again confirms that it is necessary to study mixed and componentwise condition numbers for the TLS problem. Moreover, when the TLS problem is structured, it is suitable to study the structured perturbation analysis because this will help us to understand the structured preserved algorithms; see [26]. Structured perturbation analysis for linear system, linear least squares and Tikhonov regularization problem had been investigated in [27, 28, 29, 30, 31, 32], respectively.

In this paper, under the genericity condition, we study the sensitivity of a linear function of the TLS solution  $x$  to perturbations on the data  $A$  and  $b$ , which is defined as

$$\begin{aligned}\Psi : \mathbb{R}^{m \times n} \times \mathbb{R}^m &\rightarrow \mathbb{R}^k \\ \Psi(A, b) &:= Lx,\end{aligned}\tag{1.2}$$

where  $x$  is the unique solution to the TLS problem 1.1, and  $L$  is an  $k$ -by- $n$ ,  $k \leq n$ , matrix introduced for the selection of the solution components. For example, when  $L = I_n$  ( $k = n$ ), all the  $n$  components of the solution  $x$  are equally selected. When  $L = e_i$  ( $k = 1$ ), the  $i$ th unit vector in  $\mathbb{R}^n$ , then only the  $i$ th component of the solution is selected. In the reminder of this paper, we always suppose that  $L$  is not numerically perturbed. Condition numbers for a linear function of the solution to linear system [33], linear least squares [34, 35, 36] and the TLS problem [22] had been studied extensively. Contrary to [22] for the normwise condition number of the TLS problem, in this paper, we will consider mixed and componentwise condition numbers of a linear function of the TLS solution  $x$  under unstructured and structured perturbations. For the structured TLS problem [6, 7, 8, 9, 10, 11], we consider the case that  $A$  is a linear structure matrix, for example Toeplitz matrix. Because the set  $\mathcal{S}$  of the linear structured matrix is a subspace of  $\mathbb{R}^{m \times n}$ , we assume its dimension is  $q$  and there exists a unique vector denoted by  $a$  such that

$$A = \sum_{i=1}^q a_i S_i,\tag{1.3}$$

where  $S_1, \dots, S_q$  is the basis of  $\mathcal{S}$ . In this paper, we also study the sensitivity of a linear function of the structured TLS solution  $x$  to perturbations on the data  $a$  and  $b$ , which is defined as

$$\begin{aligned}\Psi_s : \mathbb{R}^q \times \mathbb{R}^m &\rightarrow \mathbb{R}^k \\ \Psi(a, b) &:= Lx,\end{aligned}\tag{1.4}$$

where  $x$  is the unique solution to the structured TLS problem under genericity condition.

This paper is devoted to obtain the explicit expressions for mixed and componentwise condition numbers of the linear function of the TLS solution when perturbations on data are measured componentwise and the perturbations on the solution are measured either

componentwise or normwise by means of the dual techniques [35]. In particular, as also mentioned in [35], the dual techniques enable us to derive condition numbers by maximizing a linear function over a space of smaller dimension than the data space. Both the unstructured and structured condition numbers are considered. We also study the relationship between those two type condition numbers, and prove that the structured ones are smaller than the unstructured ones from the derived expressions. Moreover, the expressions of the proposed condition numbers can recover the pervious results [24, 26] when  $L = I_n$ . By taking account of the SVD method for solving TLS, we give SVD-based formulae of the proposed condition numbers. Numerical examples show that our theoretical results are effective. Especially, in Example 1 our unstructured mixed and componentwise condition numbers for a linear function of the TLS solution  $x$  can be much smaller than the nowise condition number given by [22], which means that it is more suitable to use the mixed and componentwise condition numbers to measure the conditioning of TLS when the data is sparse or badly-scaled. For the structured TLS problem, Example 3 shows it is necessary to adopt the structured mixed and componentwise condition numbers instead of using the unstructured ones to measure the conditioning for the structured TLS problem.

The paper is organized as follows. In Section 2, some basic result of the TLS problem and the dual techniques for deriving condition number [35] are reviewed. We derive the explicit expression for the unstructured and structured condition numbers, and study the relationship between them. Also we prove that our results can recover the previous condition numbers expressions of the TLS problem when  $L = I_n$ . Sharp upper bounds for unstructured mixed and componentwise condition numbers are also given. Moreover, by taking account of the SVD method for solving the TLS problem, we obtain SVD-based fomula for our proposed condition numbers. We do some numerical examples to show the effectiveness of the proposed condition numbers in Section 4. At end, in Section 5 concluding remarks are drawn.

## 2. Preliminaries

In this section we give some backgrounds on theoretical results on the TLS problem. Also, the dual techniques for deriving condition number's expressions are reviewed.

### 2.1. Basic results

Assume that we have the SVDs of  $A$  and  $[A, b]$ , respectively,

$$A = \tilde{U}\tilde{\Sigma}\tilde{V}^\top, \quad [A, b] = U\Sigma V^\top, \quad (2.1)$$

where  $U, \tilde{U} \in \mathbb{R}^{m \times m}$ ,  $\tilde{V} \in \mathbb{R}^{n \times n}$  and  $V \in \mathbb{R}^{(n+1) \times (n+1)}$  are orthogonal,  $\tilde{\Sigma} = \text{Diag}(\tilde{\sigma}_1, \tilde{\sigma}_2, \dots, \tilde{\sigma}_n) \in \mathbb{R}^{m \times n}$ ,  $\tilde{\sigma}_1 \geq \tilde{\sigma}_2 \geq \dots \geq \tilde{\sigma}_n \geq 0$  and  $\Sigma = \text{Diag}(\sigma_1, \dots, \sigma_n, \sigma_{n+1}) \in \mathbb{R}^{m \times (n+1)}$ ,  $\sigma_1 \geq \sigma_2 \geq \dots \geq \sigma_{n+1} \geq 0$ . Let  $v_{n+1}$  be the last column of  $V$  and  $v_{n+1, n+1}$  denotes its  $(n+1)$ th component of  $v_{n+1}$ . It is assumed that the *genericity* condition

$$\tilde{\sigma}_n > \sigma_{n+1}, \quad (2.2)$$

holds to ensure the existence and uniqueness of the TLS solution [1]. As mentioned in [3, Page 35], the genericity condition (2.2) is equivalent to  $\sigma_n > \sigma_{n+1}$  and  $v_{n+1,n+1} \neq 0$ . The following identities hold for the TLS solution  $x$  (cf. [1])

$$\begin{bmatrix} x \\ -1 \end{bmatrix} = -\frac{1}{v_{n+1,n+1}}v_{n+1}, \quad v_{n+1,n+1} = \frac{1}{\sqrt{1+x^\top x}}. \quad (2.3)$$

It follows from [3, Page 36, Theorem 2.7] that the TLS solution  $x$  satisfies the equation

$$(A^\top A - \sigma_{n+1}^2 I_n) x = A^\top b. \quad (2.4)$$

From the SVD of  $[A, b]$ , it is easy to check that

$$r = b - Ax = -[A, b] \begin{bmatrix} x \\ -1 \end{bmatrix} = \frac{1}{v_{n+1,n+1}}[A, b]v_{n+1} = \frac{\sigma_{n+1}}{v_{n+1,n+1}}u_{n+1}, \quad (2.5)$$

where  $u_{n+1}$  is the  $(n+1)$ th column of  $U$ .

Lemma 2.1 presents an explicit expression for the inverse of  $P$  in (2.4).

**Lemma 2.1** [25, Lemma 2] *Let  $P = A^\top A - \sigma_{n+1}^2 I_n$ . Under the genericity condition (2.2), recalling  $x$  given by (2.3), it holds that*

$$P^{-1} = Q_1 Q Q_1,$$

where  $Q = V_{11} D^{-1} V_{11}^\top$ ,  $V_{11}$  is the leading  $n \times n$  submatrix  $V$  in (2.1),  $Q_1 = I_n + x x^\top$ ,  $D = \text{Diag}(\sigma_1^2 - \sigma_{n+1}^2, \sigma_2^2 - \sigma_{n+1}^2, \dots, \sigma_n^2 - \sigma_{n+1}^2)$ .

The classical direct solver of the TLS solution  $x$  of (1.1) is to calculate the SVD of  $[A, b]$ . The detailed description of this computation is shown in [3, Algorithm 3.1]. When the full SVD of  $[A, b]$  is computed, the TLS solution  $x$  can be computed from (2.3) and  $P^{-1}$  can be computed efficiently via Lemma 2.1, which help us to derive SVD-based expressions of the proposed condition numbers.

The Fréchet derivatives of the function  $\Psi$  with respect to the input data  $[A, b]$  plays an important role in deriving condition numbers expressions, which is given in the following lemma. Let  $d\Psi([A, b])$  be the Fréchet derivative of  $\Psi$  at  $[A, b]$ .

**Lemma 2.2** [22, Proposition 1] *Under the genericity condition (2.2), the function  $\Psi$  is a continuous mapping on  $\mathbb{R}^{m \times n} \times \mathbb{R}^{m \times 1}$ . In addition,  $\Psi$  is Fréchet differentiable at  $(A, b)$  and its Fréchet derivative is given by*

$$\begin{aligned} J &:= d\Psi(A, b) \cdot (dA, db) = LP^{-1} \left[ (dA)^\top r - \left[ A^\top + \frac{2xr^\top}{1+x^\top x} \right] dAx \right] \\ &\quad + LP^{-1} \left[ A^\top + \frac{2xr^\top}{1+x^\top x} \right] db \\ &:= J_1(dA) + J_2(db), \end{aligned} \quad (2.6)$$

where  $dA \in \mathbb{R}^{m \times n}$ ,  $db \in \mathbb{R}^{m \times 1}$ .

Given the perturbations  $\Delta A$  of  $A$  and  $\Delta b$  of  $b$ . Under the genericity condition (2.2), when  $\|[\Delta A, \Delta b]\|_F$  is small enough, the perturbed TLS problem

$$\begin{aligned} & \text{minimize} && \left\| [A + \Delta A, b + \Delta b] - [\hat{A}, \hat{b}] \right\|_F \\ & \text{subject to} && \hat{b} \in \mathcal{R}(\hat{A}), [\hat{A}, \hat{b}] \in \mathbb{R}^{m \times (n+1)}, \end{aligned} \quad (2.7)$$

has a unique TLS solution  $x + \Delta x$ . The *absolute* normwise condition number [22] of  $\Psi$  is defined by

$$\text{cond}(L, A, b) = \lim_{\epsilon \rightarrow 0} \sup_{\|[\Delta A, \Delta b]\|_F \leq \epsilon} \|L\Delta x\|_2 = \max_{[\Delta A, \Delta b] \neq 0} \frac{\|L d\Psi(A, b) \cdot (dA, db)\|_2}{\|[\Delta A, \Delta b]\|_F},$$

where  $x + \Delta x$  is the TLS solution of (2.7),  $\|A\|_2$  is the spectral norm of  $A$  and the last equality is from [15]. Baboulin and Gratton [22] derived the exact SVD-based expression of  $\kappa(L, A, b)$  as follows

$$\text{cond}(L, A, b) = \sqrt{1 + \|x\|_2^2} \left\| L \tilde{V} D' [\tilde{V}^\top \ 0] V [D'' \ 0]^\top \right\|_2, \quad (2.8)$$

where

$$\begin{aligned} D' &= \text{Diag} \left( (\tilde{\sigma}_1^2 - \sigma_{n+1}^2)^{-1}, \dots, (\tilde{\sigma}_n^2 - \sigma_{n+1}^2)^{-1} \right), \\ D'' &= \text{Diag} \left( (\sigma_1^2 + \sigma_{n+1}^2)^{\frac{1}{2}}, \dots, (\sigma_n^2 + \sigma_{n+1}^2)^{\frac{1}{2}} \right). \end{aligned}$$

The relative normwise condition number corresponding to  $\text{cond}(L, A, b)$  in (2.8) can be defined by

$$\text{cond}^{\text{rel}}(L, A, b) = \lim_{\epsilon \rightarrow 0} \sup_{\|[\Delta A, \Delta b]\|_F \leq \epsilon \| [A, b] \|_F} \frac{\|L\Delta x\|_2}{\|Lx\|_2} = \frac{\text{cond}(L, A, b) \| [A, b] \|_F}{\|Lx\|_2}. \quad (2.9)$$

In the following, if  $A \in \mathbb{R}^{m \times n}$  and  $B \in \mathbb{R}^{p \times q}$ , then the *Kronecker product*  $A \otimes B \in \mathbb{R}^{mp \times nq}$  is defined by  $A \otimes B = [a_{ij} B] \in \mathbb{R}^{mp \times nq}$  [37]. Zhou et al. [24] defined and derived the relative mixed and componentwise condition numbers as follows,

$$\begin{aligned} m(A, b) &= \lim_{\epsilon \rightarrow 0} \sup_{\substack{|\Delta A| \leq \epsilon |A|, \\ |\Delta b| \leq \epsilon |b|}} \frac{\|\Delta x\|_\infty}{\|x\|_\infty} = \frac{\left\| |M + N| \begin{bmatrix} |\text{vec}(A)| \\ |b| \end{bmatrix} \right\|_\infty}{\|x\|_\infty}, \\ c(A, b) &= \lim_{\epsilon \rightarrow 0} \sup_{\substack{|\Delta A| \leq \epsilon |A|, \\ |\Delta b| \leq \epsilon |b|}} \left\| \frac{\Delta x}{x} \right\|_\infty = \left\| D_x^\dagger |M + N| \begin{bmatrix} |\text{vec}(A)| \\ |b| \end{bmatrix} \right\|_\infty, \end{aligned} \quad (2.10)$$

where we denote by  $|A| = (|a_{ij}|)$  for a given matrix  $A$ ,  $|a| \leq |b|$  represents  $|a_i| \leq |b_i|$  for two vectors  $a = [a_1, a_2, \dots, a_n]^\top$  and  $b = [b_1, b_2, \dots, b_n]^\top$ ,

$$\begin{aligned} M &= [P^{-1} \otimes b^\top - x^\top \otimes (P^{-1} A^\top) - P^{-1} \otimes (Ax)^\top \quad P^{-1} A^\top], \\ N &= 2\sigma_{n+1} P^{-1} x (v_{n+1}^\top \otimes u_{n+1}^\top), \end{aligned}$$

the notation  $\text{vec}(A)$  stacks columns of  $A$  one by one to a column vector,  $D_x^\dagger$  is the Moore-Penrose inverse [2] of the diagonal matrix  $D_x$  with  $(D_x)_{ii} = x_i$ ,  $\|\cdot\|_\infty$  is the infinity norm and the symbol  $\frac{y}{x}$  denotes the componentwise division of  $x, y$ , assuming that if  $x_i = 0$  for some index  $i$  then  $y_i$  should be zero.

The structured condition numbers for the TLS problem with linear structures were studied by Li and Jia in [26]. We first review the structured perturbation results given in [26]. Recall when  $A \in \mathcal{S}$ ,  $A$  can be determined by (1.3). Denote

$$\begin{aligned} \mathcal{M}^{st} &= [\text{vec}(S_1) \quad \cdots \quad \text{vec}(S_q)], \quad \mathcal{M}_{A,b}^{st} = \begin{bmatrix} \mathcal{M}^{st} & 0 \\ 0 & I_m \end{bmatrix}, \\ K &= P^{-1} \left( 2A^\top \frac{rr^\top}{\|r\|_2^2} G(x) - A^\top G(x) + [I_n \otimes r^\top \quad 0] \right), \quad G(x) = [x^\top \quad -1] \otimes I_m. \end{aligned} \quad (2.11)$$

The structured mixed condition number  $m_s(A, b)$  is characterized as

$$m_s(A, b) = \lim_{\epsilon \rightarrow 0} \sup_{\substack{|\Delta A| \leq \epsilon |A|, |\Delta b| \leq \epsilon |b| \\ \Delta A \in \mathcal{S}}} \frac{\|\Delta x\|_\infty}{\|x\|_\infty} = \frac{\left\| |K \mathcal{M}_{A,b}^{st}| \begin{bmatrix} |a| \\ |b| \end{bmatrix} \right\|_\infty}{\|x\|_\infty}, \quad (2.12)$$

and they also proved that  $m_s(A, b) \leq m(A, b)$ .

## 2.2. Dual techniques

Let  $\mathcal{W}$  and  $\mathcal{V}$  be the Euclidean spaces equipped scalar products  $\langle \cdot, \cdot \rangle_{\mathcal{W}}$  and  $\langle \cdot, \cdot \rangle_{\mathcal{V}}$  respectively, and we consider a linear operator  $\mathcal{L} : \mathcal{W} \rightarrow \mathcal{V}$ . We denote  $\|\cdot\|_{\mathcal{W}}$  and  $\|\cdot\|_{\mathcal{V}}$  by the corresponding norms of  $\mathcal{W}$  and  $\mathcal{V}$ , respectively. The well-known adjoint operator and dual norm are defined as follows.

**Definition 1** The adjoint operator of  $\mathcal{L}$ ,  $\mathcal{L}^* : \mathcal{V} \rightarrow \mathcal{W}$  is defined by

$$\langle \mathbf{b}, \mathcal{L}\mathbf{a} \rangle_{\mathcal{V}} = \langle \mathcal{L}^*\mathbf{b}, \mathbf{a} \rangle_{\mathcal{W}}$$

where  $\mathbf{a} \in \mathcal{W}$  and  $\mathbf{b} \in \mathcal{V}$ . The dual norm  $\|\cdot\|_{\mathcal{W}^*}$  of  $\|\cdot\|_{\mathcal{W}}$  is defined by

$$\|\mathbf{a}\|_{\mathcal{W}^*} = \max_{w \neq 0} \frac{\langle \mathbf{a}, w \rangle_{\mathcal{W}}}{\|w\|_{\mathcal{W}}}$$

and the dual norm  $\|\cdot\|_{\mathcal{V}^*}$  can be defined similarly.

Using the canonical scalar product in  $\mathbb{R}^n$ , the corresponding dual norms with respect to the common vector norms are given by :

$$\|\cdot\|_{1^*} = \|\cdot\|_\infty, \quad \|\cdot\|_{\infty^*} = \|\cdot\|_1 \quad \text{and} \quad \|\cdot\|_{2^*} = \|\cdot\|_2.$$

Let the scalar product  $\langle A, B \rangle = \text{trace}(A^\top B)$  be defined in  $\mathbb{R}^{m \times n}$ , where  $\text{trace}(A)$  is the trace of  $A$ . Then it is easy to see that  $\|A\|_{F^*} = \|A\|_F$  since  $\text{trace}(A^\top A) = \|A\|_F^2$ .



For the linear operator  $\mathcal{L}$  from  $\mathcal{W}$  to  $\mathcal{V}$ , let  $\|\mathcal{L}\|_{\mathcal{W},\mathcal{V}}$  be the operator norm induced by the norms  $\|\cdot\|_{\mathcal{W}}$  and  $\|\cdot\|_{\mathcal{V}}$ . Consequently, for linear operators from  $\mathcal{V}$  to  $\mathcal{W}$ , the norm induced from the dual norms  $\|\cdot\|_{\mathcal{W}^*}$  and  $\|\cdot\|_{\mathcal{V}^*}$ , is denoted by  $\|\cdot\|_{\mathcal{V}^*,\mathcal{W}^*}$ .

We have the following result for the adjoint operators and dual norms [35].

**Lemma 2.3** *With notations above, the following property*

$$\|\mathcal{L}\|_{\mathcal{W},\mathcal{V}} = \|\mathcal{L}^*\|_{\mathcal{V}^*,\mathcal{W}^*}$$

*holds.*

As pointed in [35], it may be desirable to compute  $\|\mathcal{L}^*\|_{\mathcal{V}^*,\mathcal{W}^*}$  instead of  $\|\mathcal{L}\|_{\mathcal{W},\mathcal{V}}$  when the dimension of the Euclidean space  $\mathcal{V}^*$  is lower than  $\mathcal{W}$  because it implies a maximization over a space of smaller dimension.

Now, we consider a product space  $\mathcal{W} = \mathcal{W}_1 \times \cdots \times \mathcal{W}_s$  where each Euclidean space  $\mathcal{W}_i$  is equipped with the scalar product  $\langle \cdot, \cdot \rangle_{\mathcal{W}_i}$  and the corresponding norm  $\|\cdot\|_{\mathcal{W}_i}$ . In  $\mathcal{W}$ , we define the following scalar product

$$\langle (\mathbf{a}_1, \dots, \mathbf{a}_s), (\mathbf{b}_1, \dots, \mathbf{b}_s) \rangle = \langle \mathbf{a}_1, \mathbf{b}_1 \rangle_{\mathcal{W}_1} + \cdots + \langle \mathbf{a}_s, \mathbf{b}_s \rangle_{\mathcal{W}_s},$$

and the corresponding product norm

$$\|(\mathbf{a}_1, \dots, \mathbf{a}_s)\|_v = v(\|\mathbf{a}_1\|_{\mathcal{W}_1}, \dots, \|\mathbf{a}_s\|_{\mathcal{W}_s}),$$

where  $v$  is an absolute norm on  $\mathbb{R}^s$ , that is  $v(|\mathbf{a}|) = v(\mathbf{a})$ , for any  $\mathbf{a} \in \mathbb{R}^s$ , see [12] for details. We denote  $v^*$  is the dual norm of  $v$  with respect to the canonical inner-product of  $\mathbb{R}^s$  and we are interested in determining the dual  $\|\cdot\|_{v^*}$  of the product norm  $\|\cdot\|_v$  with respect to the scalar product of  $\mathcal{W}$ . The following result can be found in [35].

**Lemma 2.4** *The dual of the product norm can be expressed by*

$$\|(\mathbf{a}_1, \dots, \mathbf{a}_s)\|_{v^*} = v(\|\mathbf{a}_1\|_{\mathcal{W}_1^*}, \dots, \|\mathbf{a}_s\|_{\mathcal{W}_s^*}).$$

In the following we apply adjoint operators and dual norms to derive the explicit expressions for the condition numbers of TLS. We can view the Euclidean space  $\mathcal{W}$  with norm  $\|\cdot\|_{\mathcal{W}}$  as the space of the input data in TLS and  $\mathcal{V}$  with norm  $\|\cdot\|_{\mathcal{V}}$  as the space of the solution in TLS. Then the function  $\Psi$  in (1.2) is an operator from  $\mathcal{W}$  to  $\mathcal{V}$  and the condition number is the measurement of the sensitivity of  $\Psi$  to the perturbation in its input data.

From [15], if  $\Psi$  is Fréchet differentiable in neighborhood of  $\mathbf{a} \in \mathcal{W}$ , then the absolute condition number of  $\Psi$  at  $\mathbf{a} \in \mathcal{W}$  is given by

$$\kappa(\mathbf{a}) = \|\mathbf{d}\Psi(\mathbf{a})\|_{\mathcal{W},\mathcal{V}} = \max_{\|z\|_{\mathcal{W}}=1} \|\mathbf{d}\Psi(\mathbf{a}) \cdot z\|_{\mathcal{V}},$$

where  $\|\cdot\|_{\mathcal{W},\mathcal{V}}$  is the operator norm induced by the norms  $\|\cdot\|_{\mathcal{W}}$  and  $\|\cdot\|_{\mathcal{V}}$  and  $\mathbf{d}\Psi(\mathbf{a})$  is the Fréchet derivative of  $\Psi$  at  $\mathbf{a}$ . If  $\Psi(\mathbf{a})$  is nonzero, the *relative condition number* of  $\mathbf{a}$  at  $\mathbf{a} \in \mathcal{W}$  is defined as

$$\kappa^{\text{rel}}(\mathbf{a}) = \kappa(\mathbf{a}) \frac{\|\mathbf{a}\|_{\mathcal{W}}}{\|\Psi(\mathbf{a})\|_{\mathcal{V}}}.$$



The expression of  $\kappa(\mathbf{a})$  is related to the operator norm of the linear operator  $\mathbf{d}\Psi(\mathbf{a})$ . Applying Lemma 2.3, we have the following expression of  $\kappa(\mathbf{a})$  in terms of adjoint operator and dual norm:

$$\kappa(\mathbf{a}) = \max_{\|\mathbf{da}\|_{\mathcal{W}}=1} \|\mathbf{d}\Psi(\mathbf{a}) \cdot \mathbf{da}\|_{\mathcal{V}} = \max_{\|z\|_{\mathcal{V}^*}=1} \|\mathbf{d}\Psi(\mathbf{a})^* \cdot z\|_{\mathcal{W}^*}. \quad (2.13)$$

Now we consider the componentwise metric on a data space  $\mathcal{W} = \mathbb{R}^n$ . For any given  $\mathbf{a} \in \mathcal{W}$ , the subset  $\mathcal{W}_{\mathbf{a}} \in \mathcal{W}$  is a set of all elements  $\mathbf{da} \in \mathcal{W}$  satisfying that  $\mathbf{da}_i = 0$  whenever  $\mathbf{a}_i = 0$ ,  $1 \leq i \leq n$ . Thus in a componentwise perturbation analysis, we measure the perturbation  $\mathbf{da} \in \mathcal{W}_{\mathbf{a}}$  of  $\mathbf{a}$  using the following componentwise norm with respect to  $\mathbf{a}$

$$\|\mathbf{da}\|_c = \min\{\omega, |\mathbf{da}_i| \leq \omega |\mathbf{a}_i|, i = 1, \dots, n\}. \quad (2.14)$$

Equivalently, it is easy to see that the componentwise relative norm has the following property

$$\|\mathbf{da}\|_c = \max \left\{ \frac{|\mathbf{da}_i|}{|\mathbf{a}_i|}, \mathbf{a}_i \neq 0 \right\} = \left\| \left( \frac{|\mathbf{da}_i|}{|\mathbf{a}_i|} \right) \right\|_{\infty}, \quad (2.15)$$

where  $\mathbf{da} \in \mathcal{W}_{\mathbf{a}}$ .

In the following we consider the dual norm  $\|\cdot\|_{c^*}$  of the componentwise norm  $\|\cdot\|_c$ . Let the product space  $\mathcal{W}$  be  $\mathbb{R}^n$ , each  $\mathcal{W}_i$  be  $\mathbb{R}$ , and the absolute norm  $v$  be  $\|\cdot\|_{\infty}$ . Setting the norm  $\|\mathbf{da}_i\|_{\mathcal{W}_i}$  in  $\mathcal{W}_i$  to  $|\mathbf{da}_i|/|\mathbf{a}_i|$  when  $\mathbf{a}_i \neq 0$ , from Definition 1, we have the dual norm

$$\|\mathbf{da}_i\|_{\mathcal{W}_i^*} = \max_{z \neq 0} \frac{|\mathbf{da}_i \cdot z|}{\|z\|_{\mathcal{W}_i}} = \max_{z \neq 0} \frac{|\mathbf{da}_i \cdot z|}{|z|/|\mathbf{a}_i|} = |\mathbf{da}_i| |\mathbf{a}_i|.$$

Applying Lemma 2.4 and (2.15) and recalling  $\|\cdot\|_{\infty^*} = \|\cdot\|_1$ , we derive the explicit expression of the dual norm

$$\|\mathbf{da}\|_{c^*} = \|(\|\mathbf{da}_1\|_{\mathcal{W}_1^*}, \dots, \|\mathbf{da}_n\|_{\mathcal{W}_n^*})\|_{\infty^*} = \|(|\mathbf{da}_1| |\mathbf{a}_1|, \dots, |\mathbf{da}_n| |\mathbf{a}_n|)\|_1. \quad (2.16)$$

Because of the condition  $\|\mathbf{da}\|_{\mathcal{W}} = 1$  in the condition number  $\kappa(\mathbf{a})$  in (2.13), whether  $\mathbf{da}$  is in  $\mathcal{W}_{\mathbf{a}}$  or not, the expression of the condition number  $\kappa(\mathbf{a})$  remains valid. Indeed, if  $\mathbf{da} \notin \mathcal{W}_{\mathbf{a}}$ , that is,  $\mathbf{da}_i \neq 0$  while  $\mathbf{a}_i = 0$  for some  $i$ , then  $\|\mathbf{da}\|_c = \infty$ . Consequently, such perturbation  $\mathbf{da}$  is excluded from the calculation of  $\kappa(\mathbf{a})$ . Following (2.13), we have the following lemma on the condition number in adjoint operator and dual norm.

**Lemma 2.5** *Using the above notations and the componentwise norm defined in (2.15), the condition number  $\kappa(\mathbf{a})$  can be expressed by*

$$\kappa(\mathbf{a}) = \max_{\|z\|_{\mathcal{V}^*}=1} \|(\mathbf{d}\Psi(\mathbf{a}))^* \cdot z\|_{c^*},$$

where  $\|\cdot\|_{c^*}$  is given by (2.16).

In the next section, based on Lemma 2.5, the explicit expressions for condition numbers can be deduced, where we measure the errors for the solution using componentwise perturbation analysis, while for the input data, we can measure the error either componentwise or normwise. However, regardless of the norms chosen in the solution space, we always use the componentwise norm in the data space.

### 3. Mixed and componentwise condition numbers for TLS

In this section we will derive the explicit condition numbers expressions for a linear function of the solution of TLS by means of the dual techniques under componentwise perturbations, which is introduced in [35]. Both the unstructured and structured condition number expressions are derived. Moreover, our condition numbers can recover the previous results on the mixed and componentwise condition numbers [26, 24] when we take  $L = I_n$ . Also sharp upper bounds for the unstructured mixed and componentwise condition numbers are obtained. Through using the already computed SVD for solving TLS, we can obtain SVD-based formulae for condition numbers and their upper bounds.

#### 3.1. Unstructured condition number expressions of TLS via dual techniques

In this subsection we will derive the explicit expressions of unstructured condition numbers for TLS through dual techniques stated in the previous section. Also we prove that the derived expressions and the previous ones [24] are mathematically equivalent. Sharp upper bounds absence of Kronecker product for condition numbers are given. Before that, we need the following lemma.

Using the definition of the adjoint operator and the classical definition of the scalar product in the data space  $\mathbb{R}^{m \times n} \times \mathbb{R}^{m \times 1}$ , an explicit expression of the adjoint operator of the above  $J(\mathbf{d}A, \mathbf{d}b)$  is given in the following lemma.

**Lemma 3.1** *The adjoint of operator of the Fréchet derivative  $J(\mathbf{d}A, \mathbf{d}b)$  in (2.6) is given by*

$$J^* : \mathbb{R}^k \rightarrow \mathbb{R}^{m \times n} \times \mathbb{R}^{m \times 1}$$

$$u \mapsto \left( ru^\top LP^{-1} - \left[ A^\top + \frac{2xr^\top}{1 + x^\top x} \right]^\top P^{-1} L^\top u x^\top, \left[ A^\top + \frac{2xr^\top}{1 + x^\top x} \right]^\top P^{-1} L^\top u \right).$$

**PROOF.** Using (2.6) and the definition of the scalar product in the matrix space, for any  $u \in \mathbb{R}^k$ , we have

$$\begin{aligned} \langle u, J_1(u) \rangle &= u^\top LP^{-1} \left[ (\mathbf{d}A)^\top r - \left[ A^\top + \frac{2xr^\top}{1 + x^\top x} \right] \mathbf{d}Ax \right] \\ &= \text{trace} \left( ru^\top LP^{-1} (\mathbf{d}A)^\top \right) - \text{trace} \left( xu^\top L P^{-1} \left[ A^\top + \frac{2xr^\top}{1 + x^\top x} \right] \mathbf{d}A \right) \\ &= \left\langle ru^\top LP^{-1} - \left[ A^\top + \frac{2xr^\top}{1 + x^\top x} \right]^\top P^{-1} L^\top u x^\top, \mathbf{d}A \right\rangle. \end{aligned}$$

For the second part of the adjoint of the derivative  $J$ , we have

$$\begin{aligned} \langle u, J_2(u) \rangle &= u^\top LP^{-1} \left[ A^\top + \frac{2xr^\top}{1 + x^\top x} \right] \mathbf{d}d \\ &= \left\langle \left[ A^\top + \frac{2xr^\top}{1 + x^\top x} \right]^\top P^{-1} L^\top u, \mathbf{d}b \right\rangle. \end{aligned}$$

Let

$$J_1^*(u) = ru^\top LP^{-1} - \left[ A^\top + \frac{2xr^\top}{1+x^\top x} \right]^\top P^{-1}L^\top ux^\top, \quad J_2^*(u) = \left[ A^\top + \frac{2xr^\top}{1+x^\top x} \right]^\top P^{-1}L^\top u$$

then

$$\langle J^*(u), (\mathbf{d}A, \mathbf{d}b) \rangle = \langle (J_1^*(u), J_2^*(u)), (\mathbf{d}A, \mathbf{d}b) \rangle = \langle u, J(\mathbf{d}A, \mathbf{d}b) \rangle,$$

which completes the proof.  $\square$

In fact, Lemma 3.1 establishes the same expressions of the adjoint operator of  $J$  as that in Proposition 3 of [22]. However, we use a different proof here to avoid forming the explicit Kronecker product-based matrix expression of  $J$ , which appeared in Proposition 2 of [22].

After obtaining an explicit expression of the adjoint operator of the Fréchet derivative, we now give an explicit expression of the condition number  $\kappa$  (2.13) in terms of the dual norm in the solution space in the following theorem.

**Theorem 3.1** *The condition number for the TLS problem can be expressed by*

$$\kappa = \max_{\|u\|_{\mathcal{V}^*}=1} \left\| [\mathcal{N}D_A \quad \mathcal{H}D_b]^\top L^\top \right\|_{\mathcal{V}^*,1},$$

where

$$\mathcal{N} = P^{-1} \left[ -x^\top \otimes \left( A^\top + \frac{2xr^\top}{1+x^\top x} \right) + I_n \otimes r^\top \right], \quad \mathcal{H} = P^{-1} \left[ A^\top + \frac{2xr^\top}{1+x^\top x} \right]. \quad (3.1)$$

PROOF. Let  $\mathbf{d}a_{ij}$ ,  $\mathbf{d}b_{ij}$ , be the entries of  $\mathbf{d}A$ ,  $\mathbf{d}b$  and  $\mathbf{d}d$  respectively, using (2.16), we have

$$\|(\mathbf{d}A, \mathbf{d}b)\|_{c^*} = \sum_{i,j} |\mathbf{d}a_{ij}| |a_{ij}| + \sum_{i,j} |\mathbf{d}b_{ij}| |b_{ij}|.$$

Applying Lemma 3.1, we derive that

$$\begin{aligned} \|J^*(u)\|_{c^*} &= \sum_{j=1}^n \sum_{i=1}^m |a_{ij}| \left| \left( ru^\top LP^\top - \left( A^\top + \frac{2xr^\top}{1+x^\top x} \right)^\top P^{-1}L^\top ux^\top \right)_{ij} \right| \\ &\quad + \sum_{i=1}^m |b_i| \left| \left( \left( A^\top + \frac{2xr^\top}{1+x^\top x} \right) P^{-1}L^\top u \right)_i \right| \\ &= \sum_{j=1}^n \sum_{i=1}^m |a_{ij}| \left| \left[ r_i P^{-1}e_j - x_j P^{-1} \left( A^\top + \frac{2xr^\top}{1+x^\top x} \right) e_i \right]^\top L^\top u \right| \\ &\quad + \sum_{i=1}^m |b_i| \left| \left( P^{-1} \left( A^\top + \frac{2xr^\top}{1+x^\top x} \right) e_i \right)^\top L^\top u \right| \end{aligned}$$

where  $r_i$  is the  $i$ th component of  $r$ . Then it can be verified that

$$r_i P^{-1} e_j - x_j P^{-1} \left( A^\top + \frac{2xr^\top}{1+x^\top x} \right) e_i$$

is the  $(m(j-1)+i)$ th column of the  $n \times (mn)$  matrix  $\mathcal{N}$  implying that the above expression equals to

$$\left\| \begin{bmatrix} D_A \mathcal{N}^\top L^\top u \\ D_b \mathcal{H}^\top L^\top u \end{bmatrix} \right\|_1 = \| [\mathcal{N} D_A \quad \mathcal{H} D_b]^\top L^\top u \|_1.$$

The theorem then follows from Lemma 2.5.  $\square$

The following case study discusses some commonly used norms for the norm in the solution space to obtain some specific expressions of the condition number  $\kappa$ . The proof is trivial thus is omitted.

**Corollary 3.1** *Using the above notations, when the infinity norm is chosen as the norm in the solution space  $\mathcal{V}$ , we get*

$$\kappa_\infty = \| |L\mathcal{N}| \text{vec}(|A|) + |L\mathcal{H}| |b| \|_\infty. \quad (3.2)$$

When the infinity norm is chosen as the norm in the solution space  $\mathbb{R}^n$ , the corresponding relative mixed condition number is given by

$$\kappa_\infty^{\text{rel}} = \frac{\| |L\mathcal{N}| \text{vec}(|A|) + |L\mathcal{H}| |b| \|_\infty}{\|Lx\|_\infty}. \quad (3.3)$$

In the following, we consider the 2-norm on the solution space and derive an upper bound for the corresponding condition number respect to the 2-norm on the solution space.

**Corollary 3.2** *When the 2-norm is used in the solution space, we have*

$$\kappa_2 \leq \sqrt{k} \kappa_\infty. \quad (3.4)$$

PROOF. When  $\|\cdot\|_{\mathcal{V}} = \|\cdot\|_2$ , then  $\|\cdot\|_{\mathcal{V}^*} = \|\cdot\|_2$ . From Theorem 3.1,

$$\kappa_2 = \| [\mathcal{N} D_A \quad \mathcal{H} D_b]^\top L^\top \|_{2,1}.$$

It follows from [12] that for any matrix  $B$ ,  $\|B\|_{2,1} = \max_{\|u\|_2=1} \|Bu\|_1 = \|B\hat{u}\|_1$ , where  $\hat{u} \in \mathbb{R}^k$  is a unit 2-norm vector. Applying  $\|\hat{u}\|_1 \leq \sqrt{k} \|\hat{u}\|_2$ , we get

$$\|B\|_{2,1} = \|B\hat{u}\|_1 \leq \|B\|_1 \|\hat{u}\|_1 \leq \sqrt{k} \|B\|_1.$$

Substituting the above  $B$  with  $[\mathcal{N} D_A \quad \mathcal{H} D_b]^\top L^\top$ , we have

$$\kappa_2 \leq \sqrt{k} \| [\mathcal{N} D_A \quad \mathcal{H} D_b]^\top L^\top \|_1,$$

which implies (3.4).  $\square$

By now, we have considered the various mixed condition numbers, that is, componentwise norm in the data space and the infinity norm or 2-norm in the solution space. In the rest of the subsection, we study the case of componentwise condition number, that is, componentwise norm in the solution space as well.

**Corollary 3.3** *Considering the componentwise norm defined by*

$$\|u\|_c = \min\{\omega, |u_i| \leq \omega |(Lx)_i|, i = 1, \dots, k\} = \max\{|u_i|/|(Lx)_i|, i = 1, \dots, k\}, \quad (3.5)$$

*in the solution space, we have the following expression for the componentwise condition number*

$$\kappa_c = \left\| |D_{Lx}^\dagger (|LN| \text{vec}(|A|) + |L\mathcal{H}| |b|) \right\|_\infty.$$

PROOF. The expressions immediately follow from Theorem 3.1 and Corollary 3.1.  $\square$

In the following, we will establish the equivalence relationship between  $\kappa_\infty^{\text{rel}}$ ,  $\kappa_c$  and  $m(A, b)$ ,  $c(A, b)$  respectively, when  $L = I_n$  in (1.2).

**Theorem 3.2** *When  $L = I_n$ , the expressions of  $\kappa_\infty^{\text{rel}}$  and  $\kappa_c$  are equivalent to those of  $m(A, b)$  and  $c(A, b)$  given by (2.10), respectively.*

PROOF. For  $N$  defined in (2.10), from (2.5), (2.3) and Kronecker product property, it is not difficult to see that

$$\begin{aligned} N &= 2\sigma_{n+1} P^{-1} x (v_{n+1}^\top \otimes u_{n+1}^\top) = -2P^{-1} x v_{n+1, n+1}^2 [x^\top \quad -1] \otimes r^\top \\ &= \frac{1}{1+x^\top x} P^{-1} \begin{bmatrix} -2x(x^\top \otimes r^\top) & 2xr^\top \end{bmatrix} = \frac{1}{1+x^\top x} P^{-1} [x^\top \otimes (-2xr^\top) \quad 2xr^\top]. \end{aligned}$$

For  $M$  given in (2.10), it also can be derived that

$$\begin{aligned} M &= [P^{-1} \otimes b^\top - x^\top \otimes (P^{-1} A^\top) - P^{-1} \otimes (Ax)^\top \quad P^{-1} A^\top] \\ &= [P^{-1} (I_n \otimes r^\top) - P^{-1} (x^\top \otimes A^\top) \quad P^{-1} A^\top] = P^{-1} [(I_n \otimes r^\top) - (x^\top \otimes A^\top) \quad A^\top]. \end{aligned}$$

Combing these two facts and the explicit expressions  $\kappa_\infty^{\text{rel}}$  and  $\kappa_c$  of when  $L = I_n$ , we can complete the proof of this theorem.  $\square$

The mixed and componentwise condition numbers of a linear function of the TLS solution  $x$  can recover the previous results  $m(A, b)$  and  $c(A, b)$  given by [24] when we take  $L = I_n$  in the expressions of  $\kappa_\infty^{\text{rel}}$  and  $\kappa_c$ . Also, we adopt the dual techniques to derive the condition numbers expressions, which enable us to reduce the computational complexity because the column number of the matrix expression of  $J$  is usually smaller than its row number.

By taking account of the compact form of the inverse of  $P$  given in Lemma 2.1, we can give a SVD-based formula of  $\kappa_\infty^{\text{rel}}$  and  $\kappa_c$  in the following corollary.

**Corollary 3.4** *With the notations above, we have*

$$\begin{aligned} \kappa_\infty^{\text{rel}} &= \frac{\left\| \left| LQ_1 Q Q_1 \left[ -x^\top \otimes \left( A^\top + \frac{2xr^\top}{1+x^\top x} \right) + I_n \otimes r^\top \right] \text{vec}(|A|) + \left| LQ_1 Q Q_1 \left( A^\top + \frac{2xr^\top}{1+x^\top x} \right) \right| |b| \right\|_\infty}{\|Lx\|_\infty}, \\ \kappa_c &= \left\| D_{Lx}^\dagger \left| LQ_1 Q Q_1 \left[ -x^\top \otimes \left( A^\top + \frac{2xr^\top}{1+x^\top x} \right) + I_n \otimes r^\top \right] \text{vec}(|A|) \right. \right. \\ &\quad \left. \left. + D_{Lx}^\dagger \left| LQ_1 Q Q_1 \left( A^\top + \frac{2xr^\top}{1+x^\top x} \right) \right| |b| \right\|_\infty, \end{aligned}$$

where  $Q$  and  $Q_1$  are defined in Lemma 2.1.

Although we have obtained the SVD-based expressions of  $\kappa_\infty^{\text{rel}}$  and  $\kappa_c$  in Corollary 3.9, they involve the computations of Kronecker product, which may needs extra memory to form them explicitly. In the following, we will give upper bounds for  $\kappa_\infty^{\text{rel}}$  and  $\kappa_c$  without Kronecker product. The proof of this corollary is based on Kronecker product property and the triangle inequality, and is omitted.

**Corollary 3.5** *With the notations above, denoting*

$$\begin{aligned}\kappa_\infty^{\text{U}} &= \frac{\left\| |LQ_1QQ_1|A^\top + \frac{2xr^\top}{1+x^\top x}|A||x| \right\|_\infty + \left\| |LQ_1QQ_1|A^\top|r| \right\|_\infty + \left\| \left| LQ_1QQ_1 \left( A^\top + \frac{2xr^\top}{1+x^\top x} \right) |b| \right\|_\infty}{\|Lx\|_\infty}, \\ \kappa_c^{\text{U}} &= \left\| D_{Lx}^\dagger |LQ_1QQ_1|A^\top + \frac{2xr^\top}{1+x^\top x}|A||x| \right\|_\infty + \left\| D_{Lx}^\dagger |LQ_1QQ_1|A^\top|r| \right\|_\infty \\ &\quad + \left\| D_{Lx}^\dagger \left| LQ_1QQ_1 \left( A^\top + \frac{2xr^\top}{1+x^\top x} \right) |b| \right\|_\infty,\end{aligned}$$

we have

$$\kappa_\infty^{\text{rel}} \leq \kappa_\infty^{\text{U}}, \quad \kappa_c \leq \kappa_c^{\text{U}}.$$

**Remark 1** From the numerical results of Example 1 in Section 4, the upper bounds  $\kappa_\infty^{\text{U}}$  and  $\kappa_c^{\text{U}}$  are asymptotic attainable, thus they are sharp.

### 3.2. Structured condition numbers expressions of TLS via dual techniques

In this subsection, we will focus on the structured perturbation analysis for the structured TLS problem [6, 7, 8, 9, 10, 11]. The explicit expressions are deduced and they can recover the previous results on the structured condition numbers given in [26]. Also we will prove that the structured condition numbers are smaller than the corresponding unstructured ones given in the previous subsection from their explicit expressions. We consider  $A \in \mathcal{S}$  is linear structured, i.e.,  $A = \sum_{i=1}^q a_i S_i$ , where  $S_1, \dots, S_q$  form a basis of  $\mathcal{S}$ . Let us denote  $a = [a_1, \dots, a_q]^\top$ . In view of  $\text{d}A = \sum_{i=1}^q \text{d}a_i S_i$ , and from Lemma 2.2, we can prove  $\Psi_s$  defined by (1.4) is Fréchet differentiable at  $(a, b)$  and derive its Fréchet derivative in the follow lemma.

**Lemma 3.2** *The function  $\Psi_s$  defined by (1.4) is a continuous mapping on  $\mathbb{R}^q \times \mathbb{R}^m$ . In addition,  $\Psi_s$  is Fréchet differentiable at  $(A, b)$  and its Fréchet derivative is given by*

$$\begin{aligned}J_s &:= \text{d}\Psi_s(a, b) \cdot (\text{d}a, \text{d}b) = LP^{-1}V\text{d}a + LP^{-1} \left[ A^\top + \frac{2xr^\top}{1+x^\top x} \right] \text{d}b \\ &:= J_{1s}(\text{d}a) + J_{2s}(\text{d}b),\end{aligned}\tag{3.6}$$

where  $V = [v_1, \dots, v_q] \in \mathbb{R}^{n \times q}$ ,  $v_i = S_i^\top r - \left[ A^\top + \frac{2xr^\top}{1+x^\top x} \right] S_i x$ ,  $\text{d}a \in \mathbb{R}^q$  and  $\text{d}b \in \mathbb{R}^{m \times 1}$ .

Lemma 3.3 gives the adjoint of operator of  $J_s$ . Because its proof is similar to Lemma 3.1, it is omitted here.

**Lemma 3.3** *The adjoint of operator of the Fréchet derivative  $J_s(\mathbf{d}a, \mathbf{d}b)$  in (3.6) is given by*

$$J_s^* : \mathbb{R}^k \rightarrow \mathbb{R}^q \times \mathbb{R}^m$$

$$u \mapsto \left( V^\top P^{-1} L^\top u, \quad \left[ A^\top + \frac{2xr^\top}{1+x^\top x} \right]^\top P^{-1} L^\top u \right).$$

The following theorem establishes the expressions of the structured condition number  $\kappa_s$  based on the dual techniques. We omit its proof, since it is similar to the proof of Theorem 3.1.

**Theorem 3.3** *Recalling  $\mathcal{H}$  is defined in (3.1), the condition number for the structured TLS problem can be expressed by*

$$\kappa_s = \max_{\|u\|_{\mathcal{V}^*}=1} \left\| [\mathcal{N}_s D_a \quad \mathcal{H} D_b]^\top L^\top \right\|_{\mathcal{V}^*,1},$$

where  $\mathcal{N}_s = P^{-1}V$ .

**Corollary 3.6** *Using the above notations, when the infinity norm is chosen as the norm in the solution space  $\mathcal{V}$ , we get*

$$\kappa_{s,\infty} = \left\| |L\mathcal{N}_s| |a| + |L\mathcal{H}| |b| \right\|_\infty, \quad (3.7)$$

When the infinity norm is chosen as the norm in the solution space  $\mathbb{R}^n$ , the corresponding relative structured mixed condition number is given by

$$\begin{aligned} \kappa_{s,\infty}^{\text{rel}} &= \frac{\left\| |L\mathcal{N}_s| |a| + |L\mathcal{H}| |b| \right\|_\infty}{\|Lx\|_\infty} \\ &= \frac{\left\| \sum_{i=1}^q |a_i| \left| LP^{-1} \left( \left[ A^\top + \frac{2xr^\top}{1+x^\top x} \right] S_i x - S_i^\top r \right) \right| + \left| LP^{-1} \left[ A^\top + \frac{2xr^\top}{1+x^\top x} \right] |b| \right\|_\infty}{\|Lx\|_\infty}. \end{aligned}$$

In the next theorem, we will prove that  $\kappa_{s,\infty}^{\text{rel}}$  can recover the expression of  $m_s(A, b)$  given by (2.12) when  $L = I_n$ .

**Theorem 3.4** *With the above notations, we have*

$$\kappa_{s,\infty}^{\text{rel}} = m_s(A, b),$$

when  $L = I_n$  in (1.4).



PROOF. From (2.1), (2.3), (2.5) and the fact  $v_{n+1,n+1}^2 = \frac{1}{1+x^\top x}$ , it is easy to verify that

$$\begin{aligned} A^\top r &= \frac{\sigma_{n+1}}{v_{n+1,n+1}} \left( [A, b] \begin{bmatrix} I_n \\ 0 \end{bmatrix} \right)^\top u_{n+1} = \frac{\sigma_{n+1}}{v_{n+1,n+1}} V \Sigma U^\top u_{n+1} = \frac{\sigma_{n+1}^2}{v_{n+1,n+1}} [I_n \ 0] v_{n+1} \\ &= -\frac{\|r\|_2^2}{1+x^\top x} x. \end{aligned}$$

Recalling  $K$  and  $G(x)$  given in (2.11) and using Kronecker product property, we can prove that,

$$\begin{aligned} K &= P^{-1} \left( 2A^\top \frac{rr^\top}{\|r\|_2^2} G(x) - A^\top G(x) + [I_n \otimes r^\top \ 0] \right) \\ &= P^{-1} \left( -\frac{2xr^\top}{1+xx^\top} [x^\top \otimes I_m \ -I_m] - A^\top [x^\top \otimes I_m \ -I_m] + [I_n \otimes r^\top \ 0] \right) \\ &= P^{-1} \left( [-x^\top \otimes \left( \frac{2xr^\top}{1+xx^\top} \right) \ \frac{2xr^\top}{1+xx^\top}] - [x^\top \otimes A^\top \ -A^\top] + [I_n \otimes r^\top \ 0] \right) = [\mathcal{N} \ \mathcal{H}], \end{aligned}$$

where  $\mathcal{N}$  and  $\mathcal{H}$  are defined in (3.1). It is not difficult to see that for  $V$  defined in Lemma 3.6,  $V = \begin{bmatrix} I_n \otimes r^\top & x^\top \otimes \left( A^\top \frac{2xr^\top}{1+xx^\top} \right) \end{bmatrix} \mathcal{M}^{st}$ , where  $\mathcal{M}^{st}$  is defined in (2.11). Then from (2.10),

$$\begin{aligned} m_s(A, b) &= \frac{\left\| \left| K \mathcal{M}_{A,b}^{st} \right| \begin{bmatrix} |a| \\ |b| \end{bmatrix} \right\|_\infty}{\|x\|_\infty} = \frac{\| |\mathcal{N} \mathcal{M}^{st}| |a| + |\mathcal{H}| |b| \|_\infty}{\|x\|_\infty} \\ &= \frac{\left\| \left| P^{-1} \begin{bmatrix} I_n \otimes r^\top & x^\top \otimes \left( A^\top \frac{2xr^\top}{1+xx^\top} \right) \end{bmatrix} \mathcal{M}^{st} \right| |a| + |\mathcal{H}| |b| \right\|_\infty}{\|x\|_\infty} \\ &= \frac{\| |P^{-1} V| |a| + |\mathcal{H}| |b| \|_\infty}{\|x\|_\infty} = \kappa_{s,\infty}^{\text{rel}}, \end{aligned}$$

whenever  $L = I_n$ .

□

As in the previous section, in the following corollary, we consider the 2-norm on the solution space and derive an upper bound for the corresponding structured condition number respect to the 2-norm on the solution space. The proof is similar to that of Corollary 3.2, thus we omit it.

**Corollary 3.7** *When the 2-norm is used in the solution space, we have*

$$\kappa_{s,2} \leq \sqrt{k} \kappa_{s,\infty}.$$

In Corollaries 3.6 and 3.7, we have studied the various mixed condition numbers, that is, componentwise norm in the data space and the infinity norm or 2-norm in the solution space. Again as in the previous subsection, we consider the case of componentwise condition number, that is, componentwise norm in the solution space as well.

**Corollary 3.8** *Considering the componentwise norm defined by (3.5) in the solution space, we have the following two expressions for the componentwise condition number*

$$\begin{aligned}\kappa_{s,c} &= \left\| D_{Lx}^\dagger (|L\mathcal{N}_s| |a| + |L\mathcal{H}_s| |b|) \right\|_\infty \\ &= \left\| D_{Lx}^\dagger \left( \sum_{i=1}^q |a_i| \left| LP^{-1} \left( \left[ A^\top + \frac{2xr^\top}{1+x^\top x} \right] S_i x - S_i^\top r \right) \right| \right) + \left| LP^{-1} \left[ A^\top + \frac{2xr^\top}{1+x^\top x} \right] |b| \right\|_\infty.\end{aligned}$$

In the following theorem, we will prove that the structured mixed and componentwise condition numbers are smaller than the corresponding counterparts from their derived expressions under some assumptions.

**Theorem 3.5** *Suppose that the basis  $\{S_1, S_2, \dots, S_q\}$  for  $\mathcal{S}$  satisfies  $|A| = \sum_{i=1}^q |a_i| |S_i|$  for any  $A \in \mathcal{S}$  in (1.3), then*

$$\kappa_{s,\infty}^{\text{rel}} \leq \kappa_\infty^{\text{rel}} \quad \text{and} \quad \kappa_{s,c} \leq \kappa_c.$$

PROOF. Using the monotonicity of the infinity norm, we have

$$\begin{aligned}& \left\| \sum_{i=1}^q |a_i| \left| LP^{-1} \left( \left( A^\top + \frac{2xr^\top}{1+x^\top x} \right) S_i x - S_i^\top r \right) \right| + \left| LP^{-1} \left( A^\top + \frac{2xr^\top}{1+x^\top x} \right) |b| \right\|_\infty \\ &= \left\| L \begin{bmatrix} \mathcal{N}V & \mathcal{H} \end{bmatrix} \begin{bmatrix} |a| \\ |b| \end{bmatrix} \right\|_\infty \leq \left\| \begin{bmatrix} |L\mathcal{N}| & |V| & |L\mathcal{H}| \end{bmatrix} \begin{bmatrix} |a| \\ |b| \end{bmatrix} \right\|_\infty \\ &\leq \left\| |L\mathcal{N}| \sum_{i=1}^q |a_i| |\text{vec}(S_i)| + |L\mathcal{H}| |b| \right\|_\infty = \left\| |L\mathcal{N}| \text{vec}(|A|) + |L\mathcal{H}| |b| \right\|_\infty,\end{aligned}$$

for the last equality we use the assumption  $|A| = \sum_{i=1}^q |a_i| |S_i|$ . With the above inequality, and the expressions of  $\kappa_{s,\infty}^{\text{rel}}$ ,  $\kappa_\infty^{\text{rel}}$ ,  $\kappa_{s,c}$ ,  $\kappa_c$ , it is easy to prove the first two inequalities in this theorem.  $\square$

For Toeplitz matrices, the assumption  $|A| = \sum_{i=1}^q |a_i| |S_i|$  for  $q = m + n - 1$  is satisfied, when

$$\begin{aligned}S_1 &= \text{toeplitz}(0, e_n), \dots, S_n = \text{toeplitz}(0, e_1), \\ S_{n+1} &= \text{toep}(e_2, 0) \dots, S_{m+n-1} = \text{toeplitz}(e_m, 0),\end{aligned}$$

where MATLAB's notation  $\text{toeplitz}(a, b)$  denotes a Toeplitz matrix with the first column  $a$  and first row  $b$ ,  $e_i$  is the  $i$ th column vector of a conformal dimensional identity matrix and 0 is the zero vector with a conformal dimension.

By taking account of the compact form of the inverse of  $P$  given in Lemma 2.1, we can give SVD-based formulae of  $\kappa_{s,\infty}^{\text{rel}}$  and  $\kappa_{s,c}$  in the following corollary.

**Corollary 3.9** *With the notations above, we have*

$$\kappa_{\infty}^{\text{rel}} = \frac{\left\| \sum_{i=1}^q |a_i| \left| LQ_1QQ_1 \left( \left[ A^\top + \frac{2xr^\top}{1+x^\top x} \right] S_i x - S_i^\top r \right) \right| + \left| LQ_1QQ_1 \left( A^\top + \frac{2xr^\top}{1+x^\top x} \right) \right| |b| \right\|_{\infty}}{\|Lx\|_{\infty}},$$

$$\kappa_{s,c} = \left\| D_{Lx}^\dagger \sum_{i=1}^q |a_i| \left| LQ_1QQ_1 \left( \left[ A^\top + \frac{2xr^\top}{1+x^\top x} \right] S_i x - S_i^\top r \right) \right| \right. \\ \left. + D_{Lx}^\dagger \left| LQ_1QQ_1 \left( A^\top + \frac{2xr^\top}{1+x^\top x} \right) \right| |b| \right\|_{\infty},$$

where  $Q$  and  $Q_1$  are defined in Lemma 2.1.

#### 4. Numerical examples

In this section we test some numerical examples to validate the previous derived results. All the computations are carried out using MATLAB 8.1 with the machine precision  $\mu = 2.2 \times 10^{-16}$ .

For a given TLS problem, the TLS solution is computed by (2.3). When the data  $A$  and  $b$  are generated, for the perturbations, we construct them as

$$\Delta A = 10^{-8} \cdot \Delta A_1 \odot A, \quad \Delta b = 10^{-8} \cdot \Delta b_1 \odot b, \quad (4.1)$$

where each components of  $\Delta A_1 \in \mathbb{R}^{m \times n}$  and  $\Delta b_1 \in \mathbb{R}^m$  are uniformly distributed in the interval  $(-1, 1)$ , and  $\odot$  denotes the componentwise multiplication of two conformal dimensional matrices. When the perturbations are small enough, we denote the unique solution by  $\tilde{x}$  of the perturbed TLS problem (2.7). We use the SVD method [3, Algorithm 3.1] to compute the solution  $x$  and the perturbed solution  $\tilde{x}$  via (2.3) separately.

Let  $x_{\max}$  and  $x_{\min}$  be the maximum and minimum component of  $x$  in the absolute value sense, respectively. For the  $L$  matrix in our condition numbers, we choose

$$L_0 = I_n, \quad L_1 = [e_1 \quad e_2]^\top, \quad L_2 = e_{\max}, \quad L_3 = e_{\min},$$

where max and min are the indexes corresponding to  $x_{\max}$  and  $x_{\min}$ . Thus, corresponding to the above four matrices, the whole  $x$ , the subvector  $[x_1 \quad x_2]^\top$ , the components  $x_{\max}$  and  $x_{\min}$  are selected respectively.

We measure the normwise, mixed and componentwise relative errors in  $Lx$  defined by

$$r_2^{\text{rel}} = \frac{\|L\tilde{x} - Lx\|_2}{\|Lx\|_2}, \quad r_{\infty}^{\text{rel}} = \frac{\|L\tilde{x} - Lx\|_{\infty}}{\|Lx\|_{\infty}}, \quad r_c^{\text{rel}} = \frac{\|L\tilde{x} - Lx\|_c}{\|Lx\|_c},$$

where  $\|\cdot\|_c$  is the componentwise norm defined in (3.5).

**Example 1** We construct the matrix  $A$  and vector  $b$  as follows

$$A = \begin{bmatrix} \delta & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & \delta & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \in \mathbb{R}^{9 \times 4}, \quad b = \begin{bmatrix} 1 \\ \vdots \\ 1 \end{bmatrix} \in \mathbb{R}^9,$$

where  $\delta$  is a tiny positive parameter. It is easy to see that  $A$  is sparse and badly scaled. Thus it is suitable to consider componentwise perturbation analysis for TLS.

From Table 1, it is observed that when  $\delta$  decrease from  $10^{-3}$  to  $10^{-9}$ ,  $\text{cond}^{\text{rel}}(L, A, b)$  varies from the order of  $\mathcal{O}(10^4)$  to the order of  $\mathcal{O}(10^{10})$ , while  $\kappa_{\infty}^{\text{rel}}$  and  $\kappa_c^{\text{rel}}$  are always  $\mathcal{O}(1)$ . The relative errors  $r_2^{\text{rel}}$ ,  $r_{\infty}^{\text{rel}}$  and  $r_c^{\text{rel}}$  are tiny, which means that the original TLS problem is well-conditioned. This example indicates it is more suitable to adopt  $\kappa_{\infty}^{\text{rel}}$  and  $\kappa_c^{\text{rel}}$  to measure the conditioning of the TLS problem when the data is sparse or badly-scaled. Moreover, it can be seen that the relative errors can be bounded by the asymptotic first order perturbation bounds based on the proposed condition numbers. It should be pointed out the upper bounds for  $\kappa_{\infty}^{\text{rel}}$  and  $\kappa_c^{\text{rel}}$  are asymptotic sharp, since they can be attainable from this examples. Also for different choices of  $L$ , there are differences for the relative errors and condition numbers, which tell us that it is necessary that we should consider the conditioning of the particular interested component by incorporating the matrix  $L$  in (1.2).

Table 1: Comparison of condition numbers with the corresponding relative errors for Example 1.

$\delta$	$L$	$r_2^{\text{rel}}$	$\text{cond}^{\text{rel}}(L, A, b)$	$r_\infty^{\text{rel}}$	$\kappa_\infty^{\text{rel}}$	$\kappa_\infty^{\text{U}}$	$r_c^{\text{rel}}$	$\kappa_c$	$\kappa_c^{\text{U}}$
$10^{-3}$	$I_n$	5.04e-08	1.52e+04	5.68e-08	8.43e+00	8.43e+00	6.85e-09	8.43e+00	8.43e+00
	$L_1$	5.04e-08	1.52e+04	5.68e-08	8.43e+00	8.43e+00	6.85e-09	8.43e+00	8.43e+00
	$L_2$	2.29e-08	1.64e+04	2.29e-08	2.00e+00	2.00e+00	2.29e-08	2.00e+00	2.00e+00
	$L_3$	4.29e-09	1.64e+04	4.29e-09	2.00e+00	2.00e+00	4.29e-09	2.00e+00	2.00e+00
$10^{-6}$	$I_n$	4.95e-08	1.52e+07	6.28e-08	8.43e+00	8.43e+00	1.59e-08	8.43e+00	8.43e+00
	$L_1$	4.95e-08	1.52e+07	6.28e-08	8.43e+00	8.43e+00	1.59e-08	8.43e+00	8.43e+00
	$L_2$	1.46e-08	1.64e+07	1.46e-08	2.00e+00	2.00e+00	1.46e-08	2.00e+00	2.00e+00
	$L_3$	4.46e-09	1.64e+07	4.46e-09	2.00e+00	2.00e+00	4.46e-09	2.00e+00	2.00e+00
$10^{-9}$	$I_n$	4.10e-08	1.52e+10	4.10e-08	8.43e+00	8.43e+00	8.08e-16	8.43e+00	8.43e+00
	$L_1$	4.10e-08	1.52e+10	4.10e-08	8.43e+00	8.43e+00	8.08e-16	8.43e+00	8.43e+00
	$L_2$	3.34e-07	1.64e+10	3.34e-07	2.00e+00	2.00e+00	3.34e-07	2.00e+00	2.00e+00
	$L_3$	3.23e-07	1.64e+10	3.23e-07	2.00e+00	2.00e+00	3.23e-07	2.00e+00	2.00e+00

Table 2: Comparison of condition numbers with the corresponding relative errors for Example 2.

$e_p$	$L$	$r_2^{\text{rel}}$	$\text{cond}^{\text{rel}}(L, A, b)$	$r_\infty^{\text{rel}}$	$\kappa_\infty^{\text{rel}}$	$\kappa_\infty^{\text{U}}$	$r_c^{\text{rel}}$	$\kappa_c$	$\kappa_c^{\text{U}}$
$10^0$	$I_n$	1.02e-08	1.05e+02	1.21e-08	6.37e+00	5.25e+02	5.09e-09	2.63e+01	1.41e+03
	$L_1$	7.32e-09	6.43e+01	5.67e-09	1.00e+01	1.94e+02	3.10e-09	1.08e+01	1.95e+02
	$L_2$	4.82e-09	8.24e+01	4.82e-09	1.03e+01	1.94e+02	4.82e-09	1.03e+01	1.94e+02
	$L_3$	4.72e-10	6.72e+01	4.72e-10	1.01e+01	1.94e+02	4.72e-10	1.01e+01	1.94e+02
$10^{-4}$	$I_n$	1.36e-06	6.64e+05	1.97e-06	4.84e+04	4.78e+06	6.56e-07	4.84e+04	4.78e+06
	$L_1$	1.19e-06	5.88e+05	1.20e-06	2.99e+04	2.95e+06	1.19e-06	2.99e+04	2.95e+06
	$L_2$	1.18e-06	5.88e+05	1.18e-06	2.99e+04	2.95e+06	1.18e-06	2.99e+04	2.95e+06
	$L_3$	1.19e-06	5.88e+05	1.19e-06	2.99e+04	2.95e+06	1.19e-06	2.99e+04	2.95e+06
$10^{-8}$	$I_n$	4.88e-02	6.67e+09	8.57e-02	5.15e+08	1.49e+10	6.06e-03	2.79e+09	8.07e+10
	$L_1$	1.10e-02	1.51e+09	1.10e-02	6.63e+07	1.92e+09	1.10e-02	6.63e+07	1.92e+09
	$L_2$	1.10e-02	1.51e+09	1.10e-02	6.63e+07	1.92e+09	1.10e-02	6.63e+07	1.92e+09
	$L_3$	1.10e-02	1.51e+09	1.10e-02	6.63e+07	1.92e+09	1.10e-02	6.63e+07	1.92e+09

**Example 2** We adopt the example from [22], i.e.,

$$[A, b] = Y \begin{bmatrix} D \\ 0 \end{bmatrix} Z^\top \in \mathbb{R}^{m \times (n+1)}, \quad Y = I_m - 2yy^\top, \quad Z = I_{n+1} - 2zz^\top$$

where  $y \in \mathbb{R}^m$  and  $z \in \mathbb{R}^{n+1}$  are random unit vectors and  $D = \text{Diag}(n, n-1, \dots, 1, 1-e_p)$  for a given parameter  $e_p$ . We set  $m = 100, n = 20$  as [22]. In this example, the matrix  $A$  and  $b$  are usually not spares and badly-scaled from their forms. So it cannot be argued that there should be big differences between the normwise condition numbers and mixed/componentwise condition numbers.

From the observation of Table 2, we can conclude that when  $e_p$  becomes smaller, the corresponding TLS problem tends to be more ill-conditioned. For example, when  $e_p = 10^{-8}$ , the small perturbations on  $A$  and  $b$  cause big relative errors for the TLS solution  $x$ . Because the generated data  $A$  and  $b$  is usually not sparse or badly-scaled, there are no big differences between  $\text{cond}^{\text{rel}}(L, A, b)$  and  $\kappa_\infty^{\text{rel}}(\kappa_c)$ . However, the values of  $\kappa_\infty^{\text{rel}}$  and  $\kappa_c$  are smaller to the corresponding parts of  $\text{cond}^{\text{rel}}(L, A, b)$  for different  $L$  and  $e_p$ . And the asymptotic first order perturbation bounds based on  $\kappa_\infty^{\text{rel}}$  and  $\kappa_c$  are sharper than the ones given by  $\text{cond}^{\text{rel}}(L, A, b)$ . Also, the upper bounds  $\kappa_\infty^{\text{U}}$  and  $\kappa_c^{\text{U}}$  are effective, since they are at most one hundredfold of  $\kappa_\infty^{\text{rel}}$  and  $\kappa_c$ , respectively.

**Example 3** This example is taken from [9], which is from the application in signal restoration. Let  $\alpha = 1.25$  and  $\omega = 8$ . The convolution matrix  $\bar{A}$  is an  $m \times (m-2\omega)$  Toeplitz matrix with entries in the first column given by

$$a_{i1} = \frac{1}{\sqrt{2\pi\alpha^2}} \exp \left[ \frac{-(\omega - i + 1)^2}{2\alpha^2} \right], \quad i = 1, 2, \dots, 2\omega + 1,$$

and  $a_{i1} = 0$  otherwise. The entries in the row are all zeros except  $a_{i1}$ . The target Toeplitz matrix  $A$  and right-hand side vector  $b$  are then constructed as

$$A = \bar{A} + E \text{ and } b = \bar{b} + e$$

where  $\bar{b}$  is the vector of all ones and  $E$  is a random Toeplitz matrix with the same structure as  $\bar{A}$ . The entries in  $E$  and  $e$  are generated from the standard normal distribution and scaled such that

$$\frac{\|e\|_2}{\|\bar{b}\|_2} = \frac{\|E\|_2}{\|\bar{A}\|_2} = \gamma.$$

In our test, we take  $\gamma = 0.001$  and  $m = 200$ .



Table 3: Comparison of condition numbers with the corresponding relative errors for Example 3.

$L$	$r_{\infty}^{\text{rel}}$	$\kappa_{\infty}^{\text{rel}}$	$\kappa_{s,\infty}^{\text{rel}}$	$r_c^{\text{rel}}$	$\kappa_c$	$\kappa_{s,c}$
$I_n$	5.14e-06	3.30e+04	2.50e+02	1.94e-06	4.26e+06	4.37e+03
$L_1$	6.55e-06	4.87e+04	1.69e+02	6.55e-06	4.87e+04	1.69e+02
$L_2$	6.21e-06	4.67e+04	1.75e+02	6.21e-06	4.67e+04	1.75e+02
$L_3$	5.77e-06	4.42e+04	1.80e+02	5.77e-06	4.42e+04	1.80e+02

In Table 3, the structured mixed and componentwise condition numbers are always smaller than the corresponding unstructured mixed and componentwise counterparts. The maximum and minimum ratios between the unstructured and structured ones are of  $\mathcal{O}(10^3)$  and  $\mathcal{O}(10^2)$ , respectively. So it is suitable to consider the structured perturbation analysis and measure the structured conditioning instead of the unstructured conditioning for the structured TLS problem.

## 5. Concluding Remarks

In this paper we focused on the unstructured and structured componentwise perturbation analysis for the TLS problem. Condition number expressions for the linear function of the TLS solution were derived through the dual techniques under componentwise perturbations for the input data. Moreover we studied the relationship between the new derived ones and the previous results. Sharp upper bounds for the unstructured condition numbers were given. We had proved that the structured condition numbers are smaller than the unstructured ones from the derived explicit expressions. Numerical examples validated our theoretical results.

## References

- [1] G. H. Golub, C. F. Van Loan, An analysis of the total least squares problem, SIAM J. Numer. Anal. 17 (1980) 883–893.
- [2] G. H. Golub, C. F. Van Loan, Matrix computations, Johns Hopkins Studies in the Mathematical Sciences, fourth ed., Johns Hopkins University Press, Baltimore, MD, 2013.
- [3] S. Van Huffel, J. Vandewalle, The total least squares problem, volume 9 of *Frontiers in Applied Mathematics*, Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, 1991. URL: <http://dx.doi.org/10.1137/1.9781611971002>. doi:10.1137/1.9781611971002, computational aspects and analysis, With a foreword by Gene H. Golub.
- [4] I. Markovsky, S. V. Huffel, Overview of total least-squares methods, Signal Processing 87 (2007) 2283 – 2302. Special Section: Total Least Squares and Errors-in-Variables Modeling.

- [5] Å. Björck, P. Heggernes, P. Matstoms, Methods for large scale total least squares problems, *SIAM J. Matrix Anal. Appl.* 22 (2000) 413–429 (electronic).
- [6] A. Beck, A. Ben-Tal, A global solution for the structured total least squares problem with block circulant matrices, *SIAM J. Matrix Anal. Appl.* 27 (2005) 238–255.
- [7] A. Beck, Y. C. Eldar, A. Ben-Tal, Mean-squared error estimation for linear systems with block circulant uncertainty, *SIAM J. Matrix Anal. Appl.* 29 (2007) 712–730.
- [8] N. Mastronardi, P. Lemmerling, S. Van Huffel, The structured total least squares problem, in: *Structured matrices in mathematics, computer science, and engineering, I* (Boulder, CO, 1999), volume 280 of *Contemp. Math.*, Amer. Math. Soc., Providence, RI, 2001, pp. 157–175. URL: <http://dx.doi.org/10.1090/conm/280/04627>. doi:10.1090/conm/280/04627.
- [9] J. Kamm, J. G. Nagy, A total least squares method for Toeplitz systems of equations, *BIT* 38 (1998) 560–582.
- [10] P. Lemmerling, S. Van Huffel, B. De Moor, The structured total least-squares approach for non-linearly structured matrices, *Numer. Linear Algebra Appl.* 9 (2002) 321–332.
- [11] P. Lemmerling, S. Van Huffel, Analysis of the structured total least squares problem for Hankel/Toeplitz matrices, *Numer. Algorithms* 27 (2001) 89–114.
- [12] N. J. Higham, Accuracy and stability of numerical algorithms, second ed., SIAM, Philadelphia, PA, 2002. URL: <http://dx.doi.org/10.1137/1.9780898718027>. doi:10.1137/1.9780898718027.
- [13] R. D. Fierro, J. R. Bunch, Perturbation theory for orthogonal projection methods with applications to least squares and total least squares, *Linear Algebra Appl.* 234 (1996) 71–96.
- [14] M. S. Wei, The analysis for the total least squares problem with more than one solution, *SIAM J. Matrix Anal. Appl.* 13 (1992) 746–763.
- [15] J. R. Rice, A theory of condition, *SIAM J. Numer. Anal.* 3 (1966) 287–310.
- [16] N. J. Higham, A survey of componentwise perturbation theory in numerical linear algebra, in: *Mathematics of Computation 1943–1993: a half-century of computational mathematics* (Vancouver, BC, 1993), volume 48 of *Proc. Sympos. Appl. Math.*, Amer. Math. Soc., Providence, RI, 1994, pp. 49–77. URL: <http://dx.doi.org/10.1090/psapm/048/1314843>. doi:10.1090/psapm/048/1314843.

- [17] E. Anderson, Z. Bai, C. Bischof, L. S. Blackford, J. Demmel, J. J. Dongarra, J. Du Croz, S. Hammarling, A. Greenbaum, A. McKenney, D. Sorensen, LAPACK Users' Guide (Third Ed.), Society for Industrial and Applied Mathematics, Philadelphia, PA, USA, 1999.
- [18] F. Cucker, H. Diao, Y. Wei, On mixed and componentwise condition numbers for Moore-Penrose inverse and linear least squares problems, *Math. Comp.* 76 (2007) 947–963.
- [19] I. Gohberg, I. Koltracht, Mixed, componentwise, and structured condition numbers, *SIAM J. Matrix Anal. Appl.* 14 (1993) 688–704.
- [20] J. Rohn, New condition numbers for matrices and linear systems, *Computing* 41 (1989) 167–169.
- [21] R. D. Skeel, Scaling for numerical stability in Gaussian elimination, *J. ACM* 26 (1979) 494–526.
- [22] M. Baboulin, S. Gratton, A contribution to the conditioning of the total least-squares problem, *SIAM J. Matrix Anal. Appl.* 32 (2011) 685–699.
- [23] Z. Jia, B. Li, On the condition number of the total least squares problem, *Numer. Math.* 125 (2013) 61–87.
- [24] L. Zhou, L. Lin, Y. Wei, S. Qiao, Perturbation analysis and condition numbers of scaled total least squares problems, *Numer. Algorithms* 51 (2009) 381–399.
- [25] H.-A. Diao, Y. Wei, P. Xie, Small sample statistical condition estimation for the total least squares problem, *Numer. Algor.* (2016). DOI:10.1007/s11075-016-0185-9.
- [26] B. Li, Z. Jia, Some results on condition numbers of the scaled total least squares problem, *Linear Algebra Appl.* 435 (2011) 674–686.
- [27] D. J. Higham, N. J. Higham, Structured backward error and condition of generalized eigenvalue problems, *SIAM J. Matrix Anal. Appl.* 20 (1999) 493–512 (electronic).
- [28] S. M. Rump, Structured perturbations. I. Normwise distances, *SIAM J. Matrix Anal. Appl.* 25 (2003) 1–30 (electronic).
- [29] S. M. Rump, Structured perturbations. II. Componentwise distances, *SIAM J. Matrix Anal. Appl.* 25 (2003) 31–56 (electronic).
- [30] F. Cucker, H. Diao, Mixed and componentwise condition numbers for rectangular structured matrices, *Calcolo* 44 (2007) 89–115.
- [31] H.-A. Diao, Y. Wei, S. Qiao, Structured condition numbers of structured Tikhonov regularization problem and their estimations, *J. Comput. Appl. Math.* 308 (2016) 276–300.

- [32] W. Xu, Y. Wei, S. Qiao, Condition numbers for structured least squares problems, *BIT* 46 (2006) 203–225.
- [33] S. Chandrasekaran, I. C. F. Ipsen, On the sensitivity of solution components in linear systems of equations, *SIAM J. Matrix Anal. Appl.* 16 (1995) 93–112.
- [34] M. Arioli, M. Baboulin, S. Gratton, A partial condition number for linear least squares problems, *SIAM J. Matrix Anal. Appl.* 29 (2007) 413–433.
- [35] M. Baboulin, S. Gratton, Using dual techniques to derive componentwise and mixed condition numbers for a linear function of a linear least squares solution, *BIT* 49 (2009) 3–19.
- [36] M. Baboulin, J. Dongarra, S. Gratton, J. Langou, Computing the conditioning of the components of a linear least-squares solution, *Numer. Linear Algebra Appl.* 16 (2009) 517–533.
- [37] A. Graham, *Kronecker products and matrix calculus: with applications*, Ellis Horwood Ltd., Chichester; Halsted Press [John Wiley & Sons, Inc.], New York, 1981. Ellis Horwood Series in Mathematics and its Applications.